# Reliability of hip range of motion using goniometry in pediatric femur shaft fractures

Janice Owen, BScPT;*† Derek Stephens, MSc;* James G. Wright, MD, MPH*†

**Introduction:** The purpose of this study was to determine the interrater reliability of the assessment of range of motion of the hip joint through goniometry. **Methods:** We included children aged 4 to 10 years with a femoral shaft fracture, from 4 study sites, who had had either an early hip spica cast or an external fixator. An assessor blind to treatment received at each site measured range of hip joints motion, using a standardized goniometric technique at 15 and 24 months postfracture. **Results:** The intraclass correlation coefficient (ICC) was used to quantify concordance or agreement. Most ICCs for the different aspects of hip range were between 0.2 and 0.5, indicating only slight agreement. The most reliable measure was hip flexion, with an ICC of 0.48 (95% confidence interval 0.29–0.63). **Conclusion:** Goniometric measurement, using standardized protocols for the hip, has low reliability. Only when differences in rotation exceed at least 30° and in flexion–extension exceed 50° should clinicians conclude that true change has occurred.

**Introduction :** Cette étude visait à déterminer la fiabilité entre évaluateurs d'une évaluation par goniométrie de l'amplitude du mouvement de l'articulation de la hanche. **Méthodes :** Nous avons inclus les enfants de 4 à 10 ans atteints d'une fracture du corps du fémur, provenant de quatre sites d'étude et à qui on avait posé un spica de la hanche au début ou une fixation externe. Un évaluateur ne connaissant pas le traitement reçu à chaque site a mesuré l'amplitude du mouvement des articulations de la hanche au moyen d'une technique goniométrique normalisée, 15 et 24 mois après la fracture. **Résultats :** On a utilisé le coefficient de corrélation intracatégorie (CCI) pour quantifier la concordance ou le degré d'accord. La plupart des CCI pour les différents aspects de l'amplitude du mouvement de la hanche se situaient entre 0,2 et 0,5, ce qui indique une faible concordance seulement. La mesure la plus fiable était la flexion de la hanche, avec un CCI de 0,48 (intervalle de confiance à 95 %, 0,29–0,63). **Conclusion :** La mesure goniométrique fondée sur des protocoles normalisés dans le cas de la hanche est peu fiable. C'est seulement lorsque les différences de rotation dépassent au moins 30° et la flexion en extension de 50° que les cliniciens devraient conclure qu'il y a véritablement eu changement.

Joint range of motion (ROM) is one of the objective measures used to evaluate patients' recovery after fracture.[1] Significant loss of range may contribute to impaired function. Further, loss or shifts in motion, such as hip rotation, may indicate rotational malunion. Goniometers, selected for ease of use and practical application, are usually used to measure joint range. A basic prerequisite for clinical measure is that they be reliable, which is defined as the "degree of consistency between measurements under the same conditions."[2] Clinicians need to know that clinical measures, such as the assessment of joint motion using goniometry, provide consistent results.

Goniometry has been used for both clinical and research purposes in healthy children and in special populations.[3–8] However, studies reporting on the reliability of goniometric measures have predominately focused on the adult population.[8] Furthermore, in these studies, goniometry has been undertaken in an ideal setting with cooperative subjects and static conditions. Children, in contrast, pose inherent challenges in measuring joint ROM, including their small physical size and possible lack of cooperation in maintaining an optimal position.

The purpose of this study was to examine the interrater reliability of assessment of hip range after pediatric femoral fracture.

*From the *Division of Orthopedic Surgery and †Population Health Sciences, the Hospital for Sick Children, Toronto, Ont.*

*Accepted for publication Nov. 4, 2005*

***Correspondence to:*** *Dr. James G. Wright, Hospital for Sick Children, 555 University Ave., Toronto ON M5G 1X8; fax 416 813-6433; James.Wright@sickkids.ca*

## Methods

We selected subjects from the participants (total $n$ = 101) in a randomized controlled trial comparing early hip spica and external fixator.[9] During the first 12 months after pediatric femoral shaft fractures, it was anticipated that hip ROM would initially be diminished, irrespective of treatment. However, by 1 year after the fracture, most children had regained their ultimate range.[10]

Clinicians, blind to the treatment and fracture site, evaluated children 15 and 24 months postfracture. We included all children with complete clinical assessments at the 2 time periods. Blinding was achieved by the child donning tights for the physical examination in order to cover any scars. Both the child and the parent were instructed not to identify either the fracture side or treatment with the blind assessor at each study visit. Each site ($n$ = 4) selected their own goniometer, which they used throughout the trial. Data were collected and reviewed at regular intervals during the study so that immediate clarification could be sought.

Each study site was provided with written standardized procedures for measuring active ROM of the hip and knee, as described by the American Academy of Orthopaedic Surgeons in *The Clinical Measurement of Joint Motion*.[11,12]

We determined active hip extension, using the Thomas Test to exclude any hip flexion contracture that may be present. The child was supine, and 1 hip was flexed toward the abdomen until the lumbar spine was flat. The residual lack of flexion on the opposite hip was determined as the hip flexion contracture; the arc of flexion of the opposite hip begins at this point.

We measured hip abduction and adduction with the child supine and the hips in full extension and in neutral rotation; 0° was in the midline orientation, with the 2 anterior superior iliac spines defining the axis of the pelvis. We measured hip rotation with the child prone, pelvis flat and knees flexed at 90°, allowing the legs to move apart under the force of gravity.

We used SPSS 10.1 for Windows to perform reliability, and interrater reliability was calculated with the intraclass correlation coefficient (ICC). The ICC, a test of concordance or agreement for continuous data, such as the measurement of joint range in degrees of angles, ranges from 0 to 1. An ICC greater than 0.75 indicates excellent reliability.[13] The random effects model allows for extrapolation to all raters and is useful in interrater trials.[14]

To provide a quantitative estimate of the clinical usefulness of hip range, we used limits of agreement. Limits of agreement are defined as twice the standard deviation of the differences between measures. When a difference between 2 measures exceeds the limit of agreements, differences can be more confidently attributed to true clinical change.

**Table 1**

**Description of participants by site ($n$ = 82)**

| Characteristic | Toronto | Melbourne | Los Angeles | Auckland |
|---|---|---|---|---|
| Treatment | | | | |
| Hip spica | 18 | 23 | 1 | 3 |
| External fixator | 16 | 18 | 1 | 2 |
| Side | | | | |
| Right | 17 | 27 | 1 | 5 |
| Left | 17 | 14 | 1 | 0 |
| Pattern | | | | |
| Spiral | 7 | 21 | 1 | 1 |
| Oblique | 13 | 8 | 0 | 2 |
| Transverse | 14 | 12 | 1 | 2 |
| Total | 34 | 41 | 2 | 5 |

**Table 2**

**Joint range of motion reliability ($n$ = 82)**

| | | 95% CI | |
|---|---|---|---|
| Joint range | ICC | Lower range | Upper range |
| Hip flexion | | | |
| Right | 0.48 | 0.29 | 0.63 |
| Left | 0.31 | 0.10 | 0.49 |
| Hip extension | | | |
| Right | 0.19 | –0.03 | 0.39 |
| Left | 0.22 | 0.00 | 0.41 |
| Hip abduction | | | |
| Right | 0.28 | 0.07 | 0.47 |
| Left | 0.43 | 0.24 | 0.59 |
| Hip adduction | | | |
| Right | 0.20 | –0.01 | 0.40 |
| Left | 0.19 | –0.03 | 0.39 |
| Hip internal rotation | | | |
| Right | 0.41 | 0.21 | 0.58 |
| Left | 0.30 | 0.09 | 0.49 |
| Hip external rotation | | | |
| Right | 0.06 | –0.16 | 0.27 |
| Left | 0.33 | 0.12 | 0.51 |

ICC = intraclass correlation coefficient; CI = confidence Interval.

## Results

Of the 101 trial participants, 19 had incomplete assessments and were excluded from the analysis. The average age of patients was 6 years, ranging from 4 to 10 years. We describe the patients, study sites and treatments received in Table 1.

The ICCs are presented in Table 2. All ICCs were between 0.2 and 0.5, indicating either fair or slight agreement. The highest ICC (0.48) was for right hip flexion (95% confidence interval [CI] 0.29–0.63). Scatter plots (Fig. 1, Fig. 2, Fig. 3, Fig. 4) illustrate the typical variability in interrater measurement.

The limits of agreement are provided in Table 3. Only when differences in hip rotation exceed approximately 30° can clinicians attribute that difference to true clinical change. The limits of agreement of 54° were even higher for hip flexion and extension.

## Discussion

Assessing whether change occurs as the result of disease progression or response to therapy requires a reliable measure. Measurement of hip ROM is used for many purposes. First, decreased hip motion, specifically internal rotation and abduction, are often found in many conditions affecting the hip. For example, decreased abduction is a cardinal sign of dislocated hips. Second, evaluating ROM forms part of the assessment of disease severity. For example, more severe hip arthritis is reflected in a general reduction in the hip ROM. Third, assessment of range may guide therapy. For example, restitution of motion in the early stages of Legg-Calvé-Perthes disease is a sign of treatment success. Fourth, improvement of range may be the aim of treatment. For example, the treatment aim for monoarticular arthritis may be a global increase in range. Finally, ROM can be a treatment outcome. For example, a shift in motion may indicate a rotational malunion after femoral fracture. Thus, assessment of hip motion is used often in clinical and research activities.

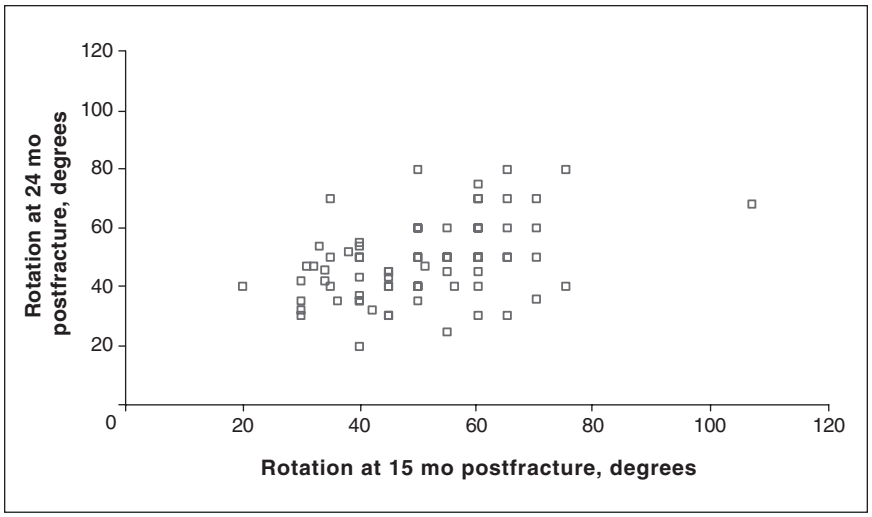


FIG. 1. Right hip internal rotation (in degrees).

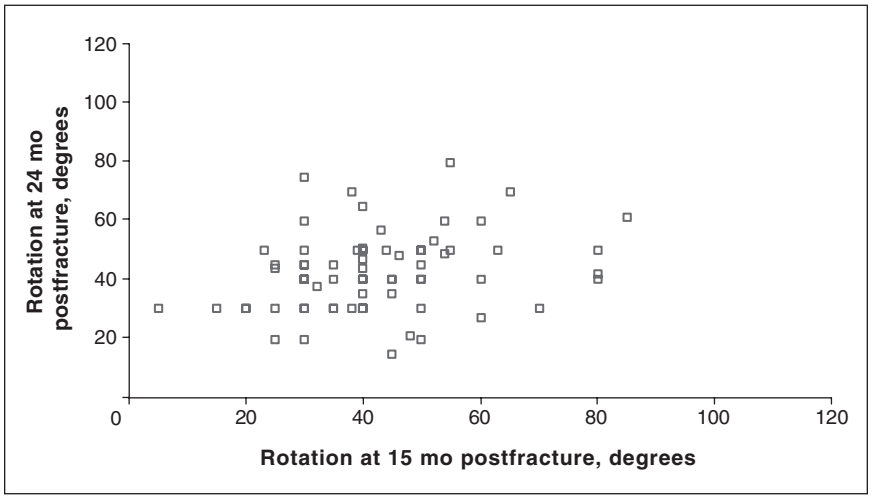


FIG. 2. Right hip external rotation (in degrees).

Only 2 studies have specifically addressed reliability of joint range in children: Watkins and colleagues[15] addressed ankle dorsiflexion and Bartlett and colleagues[16] addressed hip extension. Watkins and colleagues[15] used the Universal Goniometer and an adapted Biplane Goniometer to establish the reliability of measurements of passive ankle dorsiflexion for 25 children with either juvenile rheumatoid arthritis or cerebral palsy and 75 children with no disability. Five pairs of raters measured 5 children from each category. Standard error of measurement was calculated for each goniometer, revealing that the amount of error associated with each goniometer was similar when both rater and time were random and ranged from 6.50 for the CP group to 4.40 for the group with juvenile rheumatoid arthritis. By fixing either the occasion or the rater, or by taking the average of one rater's 2 consecutive measurements, the error was reduced by almost half.[15,16] Bartlett and colleagues[16] evaluated the reliability of 4 different methods for measuring hip fixed flexion deformity in 21 children with cerebral palsy and spina bifida. The standard deviation of 5° to 10°

was substantially less than the 14° to 21° observed in this study. However, Bartlett and colleagues[16] evaluated the reliability of only 2 experienced therapists in the context of a study. This situation is much different from our current study or from the usual clinical situation, where the variability of measurement is almost certainly much greater.

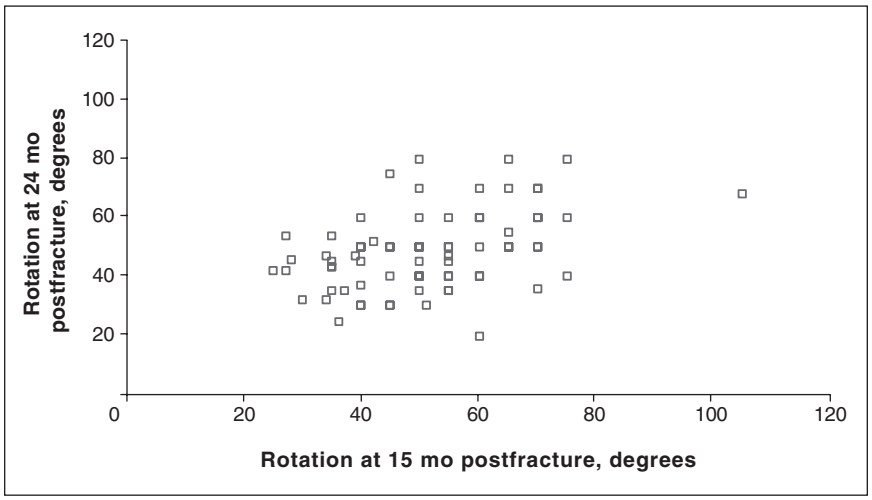


**FIG. 3. Left hip internal rotation (in degrees).**

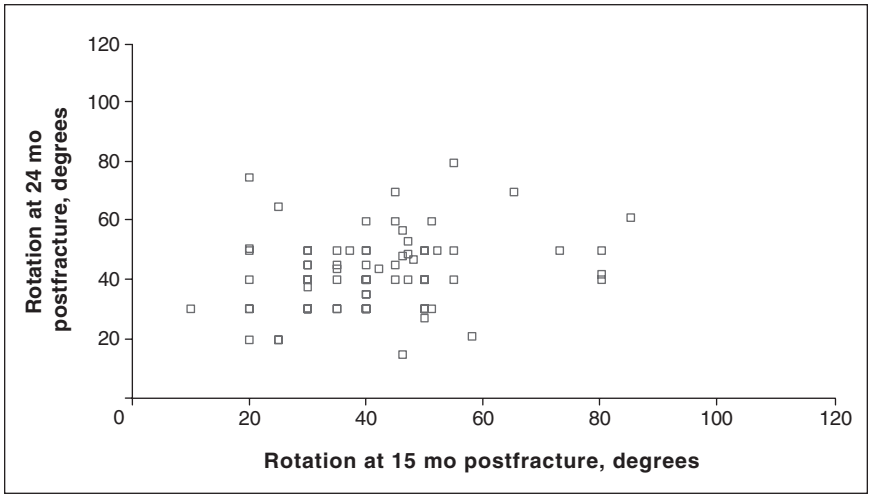


**FIG. 4. Left hip external rotation (in degrees).**

Femoral rotational malunions do not remodel with time.[17,18] In this study, we estimated only interclinician variability, and intraclinician variability is likely to be less. If anything, we probably overestimated the reliability of hip ROM assessment in the usual clinical setting. First, surgeons seldom use goniometers and often "eyeball" range assessment. Eyeball measurements are almost certainly less precise and less reliable than are goniometric measurements. Second, we used standardized written protocols, which are unlikely to exist in most clinical contexts. Third, clinical examination occurring in the context of a clinical trial is likely more reliable, compared with the usual clinical setting. Even with these caveats, we found that the reliability of hip ROM is generally low to moderate. The limits of agreement would suggest that differences of up to 30° between measurements might be due to the variability in clinical assessment. Future research will need to specifically examine and determine the sources of measurement reliability[14] in an attempt to improve the precision of the clinical assessment of hip ROM.

In conclusion, the reliability of the goniometric assessment of hip ROM is moderate at best. Only when differences in rotation exceed 30° and in flexion–extension exceed 50° can clinicians attribute the difference to clinical change.

## Table 3

**Limits of agreement for hip range of motion**

| | Right hip | | | Left hip | | |
|---|---|---|---|---|---|---|
| Movement | Difference | Mean (and SD) | 95% limits of agreement | Difference | Mean (and SD) | 95% limits of agreement |
| Flexion | –1.5 | 126.4 (19.9) | –41.3 to 38.03 | -0.07 | 125.0 (26.8) | –53.7 to 53.5 |
| Extension | 0.02 | 12.1 (14.4) | –28.8 to 28.8 | -0.07 | 12.0 (26.8) | –53.7 to 53.5 |
| Abduction | 1.8 | 46.2 (15.1) | –28.4 to 32.0 | 0.2 | 45.6 (12.1) | –24.0 to 24.4 |
| Adduction | –1.9 | 25.1 (8.9) | –19.7 to 15.9 | -2.1 | 25.5 (8.6) | –19.3 to 15.1 |
| Internal rotation | 2.3 | 49.7 (14.4) | –26.0 to 31.0 | 2.6 | 50.0 (16.0) | –29.4 to 34.6 |
| External rotation | –3.4 | 41.8 (16.1) | –33.4 to 33.4 | 1.4 | 41.5 (16.2) | –31.0 to 33.8 |

SD = standard deviation.

**Acknowledgements:** Dr. Wright is the Robert B. Salter Chair of Paediatric Surgical Research. Financial support for this study was provided by grants from the Canadian Institute of Health Research (formerly, the Medical Research Council of Canada) from 1994–2002 (Grants MT12788, MA-12788, MA-12788 renewal).

**Competing interests:** None declared.

## References

1. Bohannon R, Gajdosik R, LeVeau BF. Contribution of pelvic and lower limb motion to increases in the angle of passive straight leg raising. *Phys Ther* 1985;65: 474-6.
2. Gajdosik RL, Bohannon RW. Clinical measurement of range of motion. Review of goniometry emphasizing reliability and validity. *Phys Ther* 1987;67:1867-72.
3. Alanen JT, Levola JV, Helenius HY, et al. Ankle joint complex mobility of children 7 to 14 years old. *J Pediatr Orthop* 2001;21: 731-7.
4. Broughton NS, Brougham DI, Cole WG, et al. Reliability of radiological measurements in the assessment of the child's hip. *J Bone Joint Surg Br* 1989;71:6-8.
5. Forero N, Okamura LA, Larson MA. Normal ranges of hip motion in neonates. *J Pediatr Orthop* 1989;9:391-5.
6. Harris SR, Smith LH, Krukowski L. Goniometric reliability for a child with spastic quadriplegia. *J Pediatr Orthop* 1985;5: 348-51.
7. Katz K, Rosenthal A, Yosipovitch Z. Normal ranges of popliteal angle in children. *J Pediatr Orthop* 1992;12:229-31.
8. Kilgour GM, McNair PJ, Stott NS. Lower limb sagittal range of motion: reliability of measures and normative values. *New Zealand Journal of Physiotherapy* 2002;30: 8-24.
9. Wright JG, Wang EE, Owen JL, et al. Treatments for paediatric femoral fractures: a randomised trial. *Lancet* 2005; 365:1153-8.
10. Davids JR. Rotational deformity and remodeling after fracture of the femur in children. *Clin Orthop Relat Res* 1994;302: 27-35.
11. Greene WB, Heckman JD, editors. *The clinical measurement of joint motion*. Rosemont (IL): American Academy of Orthopaedic Surgeons; 1994. p. 152.
12. American Academy of Orthopaedic Surgeons. *Joint motion; method of measuring and recording*. Chicago: British Orthopaedic Association; 1965.
13. Fleiss J. Reliability of Measurement. In: Fleiss J, editor. *The design and analysis of clinical experiments*. New York: John Wiley and Sons; 1986. p. 1-31.
14. Wright JG, Feinstein AR. Improving the reliability of orthopaedic measurements. *J Bone Joint Surg Br* 1992;74:287-91.
15. Watkins B, Darrah J, Pain K. Reliability of passive ankle dorsiflexion measurement in children: comparison of universal and biplane goniometers. *Pediatr Phys Ther* 1995;7:3-8.
16. Bartlett MD, Wolf LS, Shurtleff DB, et al. Hip flexion contractures: a comparison of measurement methods. *Arch Phys Med Rehabil* 1985;66:620-5.
17. Davids JR. Rotational deformity and remodeling after fracture of the femur in children. *Clin Orthop Relat Res* 1994;302: 27-35.
18. Verbeek HO. Does rotation deformity, following femur shaft fracture, correct during growth? *Reconstr Surg Traumatol* 1979;17:75-81.